A Survey of *FRAXE* Allele Sizes in Three Populations

Nan Zhong, Weina Ju, Dennis Curley, Daowen Wang, James Pietrofesa, Guanyan Wu, Yan Shen, Calvin Pang, Priscilla Poon, Xixian Liu, Shen Gou, Eliisa Kajanoja, Markku Ryynänen, Carl Dobkin, and W. Ted Brown

Department of Human Genetics, New York State Institute for Basic Research for Developmental Disabilities, Staten Island, New York (N.Z., W.J., D.C., D.W., J.P., C.D., W.T.B.); China National Molecular Medical Biology Laboratory, Institute of Basic Medial Science, China Academy of Medical Science, Beijing, People's Republic of China (G.W., Y.S.); Department of Chemical Pathology, The Chinese University of Hong Kong, Shatin, Hong Kong (C.P., P.P.); Department of Medical Genetics, Tongji Medical University, Wuhan, People's Republic of China (X.L., S.G.); Division Prenatal Diagnosis, Department of Obstetrics and Gynecology, University Hospital of Kuopio, Finland (E.K., M.R.)

FRAXE **is a fragile site located at Xq27-8, which contains polymorphic triplet GCC repeats associated with a CpG island. Similar to *FRAXA*, expansion of the GCC repeats results in an abnormal methylation of the CpG island and is associated with a mild mental retardation syndrome (*FRAXE-MR*). We surveyed the GCC repeat alleles of *FRAXE* from 3 populations. A total of 665 X chromosomes including 416 from a New York Euro-American sample (259 normal and 157 with *FRAXA* mutations), 157 from a Chinese sample (144 normal and 13 *FRAXA*), and 92 from a Finnish sample (56 normal and 36 *FRAXA*) were analyzed by polymerase chain reaction. Twenty-seven alleles, ranging from 4 to 39 GCC repeats, were observed. The modal repeat number was 16 in the New York and Finnish samples and accounted for 24% of all the chromosomes tested (162/665). The modal repeat number in the Chinese sample was 18. A founder effect for *FRAXA* was suggested among the Finnish *FRAXA* samples in that 75% had the *FRAXE* 16 repeat allele versus only 30% of controls. Sequencing of the *FRAXE* region showed no imperfections within the GCC repeat region, such as those commonly seen in *FRAXA*. The smaller size and limited range of repeats and the lack of imperfections suggests the molecular mechanisms underlying *FRAXE* triplet mutations may be different from those underlying *FRAXA*.**

KEY WORDS: ***FRAXE*, polymorphism, triplet repeat, sequence, population genetics**

INTRODUCTION

FRAXE is a fragile site associated with mental retardation located in Xq2.7-8 [Sutherland and Baker, 1992; Flynn et al., 1993], approximately 600 kb distal to *FRAXA*, the fragile X syndrome fragile site. The *FRAXE* mutation is an expansion of a GCC repeat that results in abnormal methylation of a nearby CpG island located approximately 800 bp proximal to it and which has been observed in some mildly retarded individuals [Knight et al., 1993, 1994; Hamel et al., 1994; Mulley et al., 1995]. The expanded GCC size in such individuals ranges from 350 to 2,600 bp. Abnormal methylation in the GCC-associated CpG island occurs when the triplet repeat number exceeds 200 [Knight et al., 1993]. The *FRAXE* gene has not been characterized yet, although candidate transcripts have been recently identified [Gedeon et al., 1995]. The *FRAXE*-associated mental retardation (*FRAXE-MR*) phenotype is quite variable but tends to include slow learning, hyperactivity, and language delay [Flynn et al., 1993; Knight et al., 1993, 1994; Hamel et al., 1994; Mulley et al., 1995].

The molecular mechanisms which underlie the *FRAXE* GCC repeat expansion are not yet clear. In *FRAXA*, surveys of repeat sequences indicate that the mutational repeat expansion occurs within the 3′ CGG repeat region [Eichler et al., 1994; Hirst et al., 1994; Kunst and Warren, 1994; Snow et al., 1994; Zhong et al., 1995] and that it may result from the loss of AGG triplet interruptions within the CGG repeat. No comparable surveys of *FRAXE* GCC repeats are currently available. In addition, although apparent founder chromosome effects with flanking microsatellites and *FRAXA* mutations have been studied [Richards et al.,

Received for publication September 25, 1995; revision received December 15, 1995.

Address reprint requests to W. Ted Brown, Department of Human Genetics, New York State Institute for Basic Research for Developmental Disabilities, Staten Island, NY 10314.

1992; Oudet et al., 1993; Zhong et al., 1994a,b, 1996a], there are no such studies of possible founder effects for *FRAXE* mutations. However, only a few *FRAXE* mutation families have been identified [Knight et al., 1993, 1994; Mulley et al., 1995].

To investigate whether the mutational mechanisms underlying expansion in *FRAXE* may be similar to *FRAXA*, we analyzed the *FRAXE* repeat distributions in 3 populations. We screened 665 chromosomes in populations that included normal controls and *FRAXA* mutations. Repeat size modes and ranges were smaller than those for *FRAXA*. We sequenced a representative sample of *FRAXE* alleles and found no "imperfections" interrupting the GCC repeats. We found repeat distribution differences among different populations. Linkage disequilibrium between *FRAXA* and *FRAXE* alleles was found only in the Finnish population. The smaller size and limited range of repeats and the lack of imperfections suggest that the molecular mechanisms underlying *FRAXE* triplet mutations may be different from those underlying *FRAXA*.

MATERIALS AND METHODS

Population Samples

Three sets of DNA samples representative of Euro-American, Chinese, and Finnish populations, which included 665 samples that were used previously as controls (459) and subjects (206) for *FRAXA* studies, were examined at the *FRAXE* locus. The Euro-American DNA samples were obtained from unrelated individuals from the New York area, including approximately 40% who had Ashkenazi Jewish background [Brown et al., 1993; Zhong et al., 1993]. The Chinese population samples came from the southern area around Hong Kong, the middle area around Wuhan [Zhong et al., 1994b], and the northern area around Beijing. The Finnish DNA samples came from Eastern Finland [Zhong et al., 1996a].

Amplification by Polymerase Chain Reaction (PCR)

The *FRAXE* triplet repeat sizes were determined with PCR [Knight et al., 1993] by using a modified method. Twenty to 50 ng of genomic DNA was amplified in a 10-μl reaction volume containing 10 mM Tris, 50 mM KCl, 10% DMSO, 0.01% gelatin, 1.5 mM dNTPs with 75% of 7-deaza-dGTP, 1Ci (α^{32}P-dCTP, 25 pmol of each primer [603: 5′-CCT GTG AGT GTG TAA GTG TGT GAT GCT GCC G; 598: 5′-GCG AGG AAG CGG CGG CAG TGG CAC TGG G) and 2.5 units of Taq polymerase (Perkin-Elmer). The amplifications were carried out in PE9600 (Perkin-Elmer) or PTC-100 (MJ Research) thermocyclers with 9 min of denaturation at 95°C followed by 30 cycles of 1 min of denaturation at 95°C, 1 min of annealing at 65°C, and 1.5-min extension at 72°C. The final extension was 10 min. Then, 5 μl of formamide loading buffer was added to the PCR reaction tube and 4-μl aliquots were loaded onto Sequagel-6 (National Diagnostics, Atlanta, GA). After 5 hr of electrophoresis at 2,000 V, the gel was fixed for 5 min by spreading 100 ml of 15% acetic acid/methanol onto the gel surface. The gel was dried at 80°C for 45 min and exposed to Kodak XAR-5 film at −70°C. The CGG repeat of *FMR1* was sized as described elsewhere [Brown et al., 1993; Zhong et al., 1993].

Cycle Sequencing of GCC Repeat of *FRAXE* Region

The GCC region of the *FRAXE* was sequenced with primers 603 and 598 from both directions by cycle sequencing of a PCR product that represented GCC repeat region. The PCR reaction volume was 50 μl, and the PCR product was isolated by 4% Nusieve agarose (FMC) electrophoresis. It was purified with a Magic PCR Prep Kit (Promega) to yield a final elution volume of 75 μl, and 5–10 μl was used in the sequence reaction. A dsDNA cycle sequencing kit (BRL) was used according to the supplied instructions, which were modified with a final 10% DMSO (Sigma) in the cycle sequencing reaction. The Taq polymerase from the kit was replaced by AmpliTaq (Perkin-Elmer) because we found that the AmpliTaq is less inhibited by 10% DMSO.

RESULTS

Distribution of *FRAXE* GCC Repeats

A total of 27 different alleles (Fig. 1), ranging from 4 to 39 GCC repeats, was identified among the 665 chromosomes examined. The most common allele size was 16 repeats, which accounted for 24% of the chromosomes, and the overall heterozygosity was 87.7%. The *FRAXE* allele distributions and heterozygosities in the 3 populations are given in Table I.

The population repeat distributions were unimodal but showed some differences, as illustrated in Fig. 2. In the New York sample, the modal allele in normal and *FRAXA* samples was 16 repeats (27% and 24.8%, respectively). In the Chinese sample, however, the modal allele was 18 repeats (23.6%) in the controls and 17 repeats (46.2%) in the *FRAXA* samples. In the Finnish sample, 30.4% of the normal controls had the 16 repeat allele, which is similar to the New York control sample, but 75% of the Finnish *FRAXA* chromosomes had the 16 repeat allele, and only 5 different alleles were observed.

We sequenced 21 different *FRAXE* GCC alleles, ranging from 4 to 34 repeats, from the New York population. Typical results are illustrated in Fig. 3. The sequence of an allele (Fig. 3, lane C) with 34 repeats is shown in Fig. 4. None of the alleles sequenced showed any imperfections in the GCC repeat region such as the AGGs commonly seen in *FRAXA* alleles.

DISCUSSION

FRAXE-MR is 1 of 8 triplet repeat diseases identified so far [Warren and Ashley, 1995], including the fragile X syndrome (*FRAXA*), Kennedy disease (Spinal bulbar muscular atrophy, or SBMA), myotonic dystrophy (DM), Huntington disease (HD), spinocerebellar ataxia 1 (SCA1), dentatorubral–pallidoluysian atrophy (DRPLA), and Machado–Joseph disease (MJD). The prevalence of *FRAXE-MR* appears to be much lower than that of fragile X syndrome (*FRAXA*) because few cases have been identified [Mulley et al., 1995]. We have identified only 2 families with *FRAXE* mutations [unpublished data] and a screening of 300 males who were developmentally delayed identified none [Allingham-Hawkins and Ray, 1995]. The lower prevalence may reflect re-

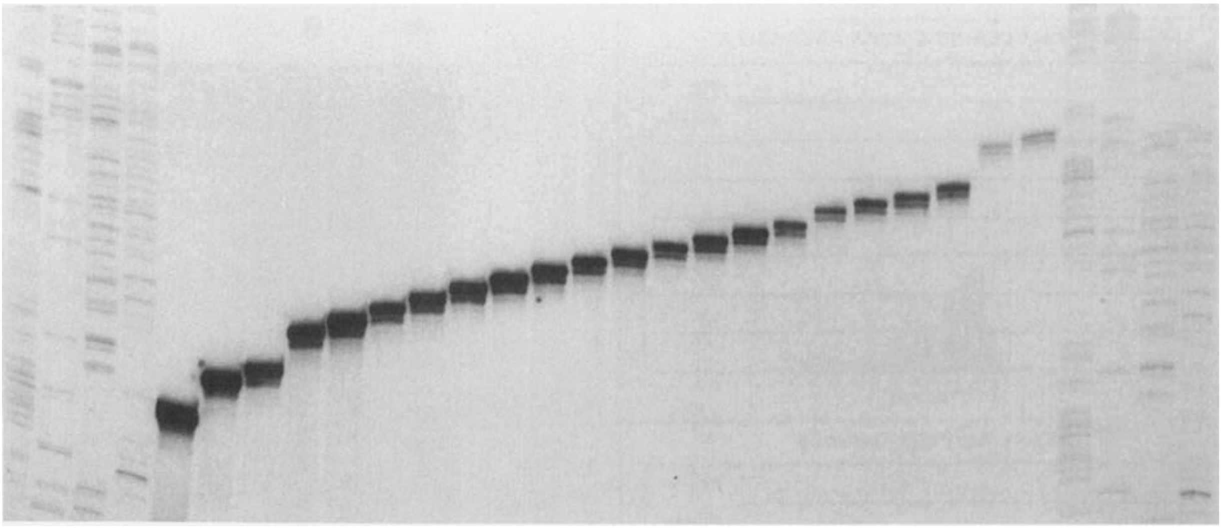

Fig. 1. PCR amplification of the *FRAXE* GCC repeat region. The GCC repeats ranged from 4 to 34 triplets. Also identified was a 39 repeat allele that is not shown. pUC18 sequence served as a size marker.

TABLE I. Distribution of *FRAXE* GCC Repeats in Three Populations (Percent of Sample)[a]

Allele[b] size (GCC repeats)	Normal			Fra(X)			Total	
	New York	Chinese	Finnish	New York	Chinese	Finnish	Normal	Fra(X)
n	259	144	56	157	13	36	459	206
4			0.6				0.5	
7			1.8				0.2	
8	0.77						0.4	
9	0.39	0.7					0.4	
10		2.0					0.6	
11	0.39	1.4					0.6	
12	0.39	4.2	1.8	0.6			1.7	0.5
13	0.77			1.2			0.4	1.0
14	0.77	4.9	1.8	1.2			2.2	1.0
15	0.92	4.2	19.6	5.7		8.3	8.9	5.8
16	**27.0**	4.2	**30.4**	**24.8**	23.1	**75.0**[c]	**20.3**	**33.5**[e]
17	13.5	13.2	8.9	18.5	**46.2**[d]	11.1	12.9	18.9
18	7.7	**23.6**	3.6	11.5	23.1		12.2	10.2
19	12.0	16.3	16.1	6.3	7.7		13.5	5.3
20	11.2	10.4	7.1	5.1		2.8	10.5	4.4
21	5.0	5.6		3.8			4.6	2.9
22	3.9	3.5	1.8	2.5			3.5	1.9
23	2.3	2.0	1.8	6.4			2.2	4.8
24	1.9	1.4	1.8	2.5			1.7	1.9
25		1.4	3.6	5.1		2.8	0.8	4.4
26	0.39	1.4		1.2			0.6	1.0
27	0.39			0.6			0.2	0.5
28	1.1						0.6	
29				0.6				0.5
30	0.39						0.2	
34	0.39			0.6			0.2	0.5
39				0.6				0.5
Heterozygosity	86.2	87.6	82.8	87.1	67.4	41.7	88.5	82.7

[a] Modes are shown in boldface type.
[b] Alleles = 27.
[c] Finnish fragile X versus normal controls, $\chi^2 = 15$, $P < 0.001$.
[d] Chinese fragile X versus normal controls, $\chi^2 = 7.4$, $P < 0.01$.
[e] General distributions of fragile X versus normal controls, $\chi^2 = 12.8$, $P < 0.001$.

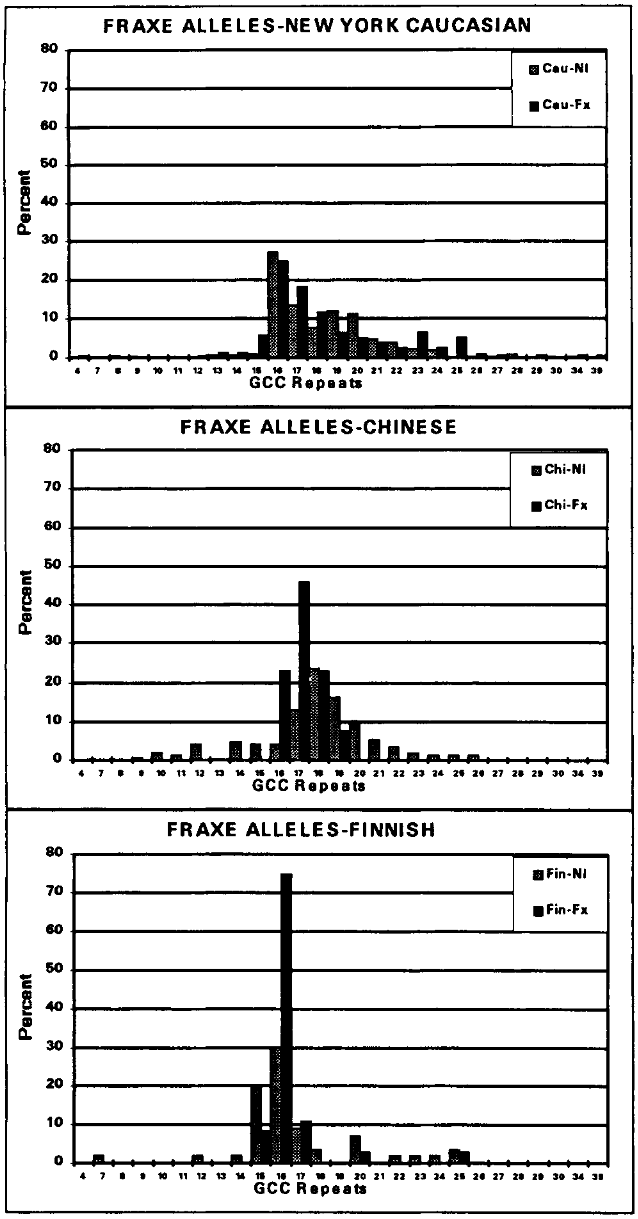

Fig. 2. Distribution of *FRAXE* GCC repeat in 3 populations. The modal allele in the New York and Finnish samples was 16 repeats in both normal control (Cau-Nl and Fin-Nl) and *FRAXA* (Cau-Fx and Fin-F_x) chromosomes. However, it was 18 in the normal (Chi-Nl) and 17 in the *FRAXA* (Chi-Fx) chromosomes in the Chinese population. X axis is the GCC repeat size, Y axis is the frequency (%) of allele distribution.

duced subject identification due to the relative mildness of the phenotype or to a reduced incidence of the mutation. This reduction in mutation prevalence may be related to the lower average and range of *FRAXE* repeat numbers. In this regard, there is little information about the normal distribution of *FRAXE* alleles [Knight et al., 1993, 1994; Allingham-Hawkins and Ray, 1995].

Our survey of 3 populations showed that the *FRAXE* GCC repeat alleles have a broad but unimodal distribution. The broad distribution is similar to that of *FMR1* alleles observed in various populations [Zhong et al., 1994a,b], but the modal *FRAXE* repeat number and range are considerably reduced. The modal *FRAXE* repeat number of 16 in the New York and Finnish samples (Table I) was in agreement with a Canadian study of 300 individuals [Allingham-Hawkins and Ray, 1995]. It differs from the mode of 15 reported by Knight et al. [1994] from a screening of 86 British subjects, which suggests either laboratory or ethnic variation. The Chinese control mode of 18 adds support to the conclusion that ethnic differences exist in triplet repeat allele distributions. The range of allele sizes observed in the 3 studies were similar: 6–25 [Knight et al., 1994], 7–35 [Allingham-Hawkins and Ray, 1995], and 4–39 [this study].

One hypothesis for the mechanism leading to the fragile X mutation is that loss of anchoring AGGs and conversion to a pure CGG repeat sequence increase the probability of expansion [Richards and Sutherland, 1992]. Observations that the instability of premutation *FMR1* alleles occurs within the 3′ pure CGG repeat re-

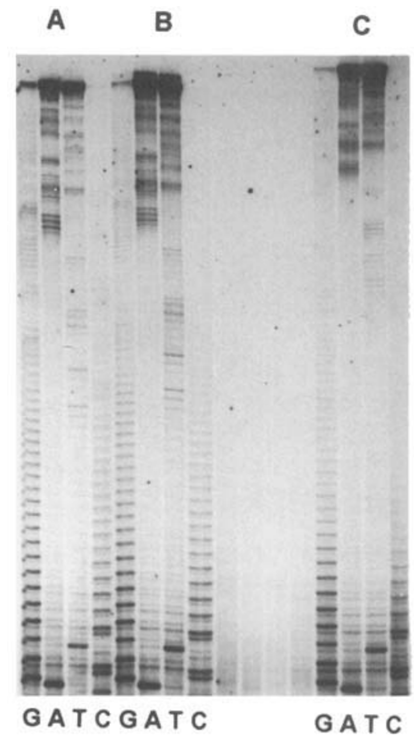

Fig. 3. Double-strand cycle sequencing of *FRAXE* GCC repeats. **Lane A:** 7. **Lane B:** 19. **Lane C:** 34 with primer 603 (forward). Only pure GCC repeats without interruption were detected.

```
5'-CCTGTGAGTG  TGTAAGTGTG  TGATGCTGCC  GCCGCCGCCG  CCGCCTGTGC
AGCCGCT(GCC)34 GCTGCCGCCC  CGGCTGCCGC  GCCGCGCCGC  TGCCTCTGCC
CCCGGCCGCC  CCCGCCGCCG  CTGCCGCCGC  CGGCCCGCAG  CCAGCCATGC
GGGCGGCCCA  GCCCGCCTGA  GCCACTCGCA  GCTGCCGCCG  CAGCGTCGGG
TCGCTGGGTG  CGCGGGCTAC  CGCGGACCGA  GCGGACCCGA  GTGGGCGACC
AGGCGCTTGC  CCGCCCAGTG  CCACTGCCGC  CGCTTCCTCG  C-3
```

Fig. 4. Sequence of *FRAXE* “C” allele (Fig. 3) of 34 repeats. The locations of the 5′ primer 603 (forward) and the 3′ primer 598 (reverse) are underlined.

gion supports this hypothesis [Eichler et al., 1994; Zhong et al., 1995b]. However, because sequencing of control *FRAXE* alleles did not reveal any interruptions within the GCC repeat region (Fig. 3), such a loss appears to be an unlikely mechanism for generating mutations of this locus. Therefore, the mechanism is likely to be similar to that of other triplet repeat diseases, such as HD and DM, in which interruptions are not found and where the threshold for instability is approximately 40 repeats. In this regard, the largest *FRAXE* alleles we observed were 34 among controls and 39 among fragile X chromosomes (Table I).

A 4-allele model was proposed for the generation of *FRAXA* mutations, in which the *FRAXA* CGG expansion was explained as a series of transitions from normal stable small alleles, to intermediate "gray-zone" alleles, to premutation alleles, and to full mutation alleles [Morton and Macpherson, 1992]. This model was expanded to a 10-allele model [Kolehmainen, 1994] and then to an n-allele model [Morris et al., 1995]. We found that gray-zone alleles of 40–45 CGG repeats for *FMR1* [Verkerk et al., 1991] are associated with larger alleles of flanking microsatellites [Zhong et al., 1996b]. If the threshold of about 40 repeats holds for *FRAXE* as for the other uninterrupted triplet repeat diseases, then *FRAXE* expansions may conform to such models.

There was no difference in frequency for the modal repeat number of 16 between the normal (27%) and fragile X (24.8%) in the New York population. However, in the Finnish population, 75% of fragile X chromosomes had the *FRAXE* 16 repeat allele versus 30.4% of the normal controls, a highly significant difference. The relation of *FMR1* founder haplotypes to *FRAXE* alleles in the Finnish population are further analyzed in a related study [Zhong et al., 1996a]. A significant difference was also seen in the Chinese population, where 46.2% of 13 fragile X chromosomes had the 17 repeat allele versus 13.2% of 144 control alleles. Although the fragile X sample size was small and this difference could reflect a sampling variation, these results do suggest that founder effects for *FMR1* mutations and *FRAXE* repeat alleles may exist in the Chinese population as well.

ACKNOWLEDGMENTS

This work was supported in part by grant HD 29407 from the NIH.

REFERENCES

Allingham-Hawkins DJ and Ray PN (1995): *FRAXE* expression is not a common etiological factor among developmentally delayed males. Am J Hum Genet 56:72–76.

Brown WT, Houck GE Jr, Jeziorowska A, Levinson FN, Ding X, Dobkin C, Zhong N, Henderson J, Sklower B, Jenkins EC (1993): Rapid fragile X carrier screening and prenatal diagnosis by a nonradioactive PCR test. JAMA 270:1569–1575.

Eichler EE, Holden J, popovich BW, Reiss AL, Snow K, Thibodeau SN, Richards DS, Wars PA, Nelson DL (1994): Length of uninterrupted CGG repeats determines instability in the *FMR*1 gene. Nature Genet 8:89–94.

Flynn GA, Hirst MC, Knight SJL, Macpherson JN, Barber JCK, Flannery AV, Davis KE, Buckle VJ (1993): Identification of the *FRAXE* fragile site in two families ascertained for X linked mental retardation. J Med Genet 30:97–100.

Gedeon AK, Keinanen M, Ades LC, Kaariainen H, Baker JGE, Sutherland GR, Mulley JC (1995): Overlapping submicroscopic deletions in Xq28 in two unrelated boys with developmental disorders: Identification of a gene near *FRAXE*. Am J Hum Genet 56:907–914.

Hamel BCJ, Smits APT, Graaff ED, Smeets DFCM, Schoute F, Eussen BHJ, Knight SJL,Davis KE, Assman-Hulmans CFCH, Oostra BA (1994): Segregation of *FRAXE* in a large family: Clinical, psychometric, cytogenetic, and molecular data. Am J Hum Genet 55:923–931.

Hirst MC, Grewal PK, Davis KE (1994): Precursor arrays for triplet expansion at the fragile X locus. Hum Molec Genet 3:1553–1560.

Knight SJL, Flannery AV, Hirst MC, Campbell, Christodoulou Z, Phelps SR, Pointon J, Middleton-Price HR, Barnicoat A, Pembrey ME, Holland J, Oostra BA, Bobrow M, Davis KE (1993): Trinucleotide repeat amplification and hypermethylation of a CpG island in *FRAXE* mental retardation. Cell 74:127–134.

Knight SJL, Voelckel MA, Hirst MC, Flannery AV, Moncla A, Davis KE (1994): Triplet repeat expansion at the *FRAXE* locus and X-linked mild mental handicap. Am J Hum Genet 55:81–86.

Kolehmainen K (1994): Population genetics of fragile X: A multiple allele model with variable risk of CGG repeat expansion. Am J Med Genet 51:428–435.

Kunst CB, Warren ST (1994): Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. Cell 77:853–861.

Morris A, Morton NE, Collins A, Macpherson J, Nelson D, Sherman S (1995): An n-allele model for progresive amplification in the *FMR*1 locus. Proc Natl Acad Sci USA 92:4833–4837.

Morton NE, Macpherson JN (1992): Population genetics of the fragile X syndrome: Multiallelic model for the *FMR*1 locus. Proc Natl Acad Sci USA 89:4215–4217.

Mulley JC, Yu S, Loesch DZ, Hay DA, Donnelly A, Gedeon AK, Carbonell P, Lopez I, Glover G, Gabarron I, Baker E, Haan EA, Hockey A, Knight SJL, Davis KE, Richards R, Sutherland GR (1995): *FRAXE* and mental retardation. J Med Genet 32:162–169.

Oudet C, Mornet E, Serre L, Thomas F, Lentes-Zengeling S, Kretz C, Deluchat C, Tejada I, Boue J, Boue A, Mandel JL (1993): Linkage disequilibrium between the fragile X mutation and two closely linked CA repeats suggests that fragile X chromosomes are derived from a small number of founder chromosomes. Am J Hum Genet 52:297–304.

Richards RI, Sutherland GR (1992): Dynamic mutations: A new class of mutations causing human disease. Cell 70:709–712.

Richards RI, Holman K, Friend K, Kremer E, Hillen D, Staples A, Brown WT, Goonewardena P, Tarleton J, Schwartz C, Sutherland GR (1992): Evidence of founder chromosomes in fragile X syndrome. Nature Genet 1:257–260.

Snow K, Tester DJ, Krucheberg KE, Schaid DJ, Thibodeaux SN (1994): Sequence analysis of the fragile X trinucleotide repeat: Implications for the origin of the fragile X mutation. Hum Molec Genet 3:1543–1551.

Sutherland GR, Baker E (1992): Characterisation of a new rare fragile site easily confused with the fragile X. Hum Molec Genet 1:111–113.

Verkerk AJMH, Pieretti M, Sutcliffe JS, Fu YH, Kohl DPA, Pizzuti A, Reiner O, Richards S, Victoria MF, Zhang F, Eussen BE, van Ommen GJB, Blonden LAJ, Riggins GJ, Chastain JL, Kunst CB, Galjaard H, Caskey CT, Nelson DL, Oostra BA, Warren ST (1991): Identification of a gene (*FMR*1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell 65:905–914.

Warren ST, Ashley CT (1995): Triplet repeat expansion mutations: The example of fragile X syndrome. Annu Rev Neurosci 18:77–99.

Zhong N, Dobkin C, Brown WT (1993): A complex mutable polymorphism located within the fragile X gene. Nature Genet 5:248–253.

Zhong N, Ye LL, Dobkin C, Brown WT (1994a): Fragile X founder chromosome effects: Linkage disequilibrium or microsatellite heterogeneity? Am J Med Genet 51:405–411.

Zhong N, Liu X, Gou S, Houck GE, Li S, Dobkin C, Brown WT (1994b): Distribution of *FMR*1 and associated microsatellite alleles in a normal Chinese population. Am J Med Genet 51:417–422.

Zhong N, Yang WH, Dobkin C, Brown WT (1995): Fragile X gene instability: Anchoring AGGs and linked microsatellites. Am J Hum Genet 57:351–361.

Zhong N, Kajanoja E, Smits B, Pietrofesa J, Curley D, Wang D, Ju W, Nolin S, Dobkins C, Ryynanen M, Brown WT (1996a): Fragile X founder effects and new mutations in Finland. Am J Med Genet 64: 226–333.

Zhong N, Ju W, Pietrofesa J, Wang D, Dobkin C, Brown WT (1996b): Fragile X "gray-zone" alleles: AGG patterns, expansion risks, and associated haplotypes. Am J Med Genet 64:261–265.